(12) UK Patent Application (19) GB (11) 2 161 004 A

(43) Application published 2 Jan 1986

(21) Application No 8409539

(22) Date of filing 12 Apr 1984

(51) INT CL⁴
H03M 11/00

(52) Domestic classification
G4H KJ
U1S 2123 2213 2215 2216 2247 G4H

(56) Documents cited
GB A 2118749 GB A 2071018 GB A 2062916
GB A 2100899 GB A 2066534 GB A 2057973

(58) Field of search
G4H

(71) Applicants
Jin-Kai Li,
No 140 5th floor, Beijing Normal University, Beijing, China.
Kam-Fu Wong,
22nd floor, Wu Sang House, 655 Nathan Road, Kowloon, Hong Kong.
Ping Lu,
9 Dai Fhun Street, Pai Po Lnb Eftate, New Territories, Hong Kong

(72) Inventor
Jin-Kai Li

(74) Agent and/or Address for Service
Lloyd Wise Tregear & Co,
Norman House,105–109 Strand, London WC2R 0AE

(54) System for encoding characters

(57) A system for encoding all kinds of characters in the world in digits is characterized by: formulating certain basic categories of strokes, to each of which allocating a digit in graphic (positional) order, namely, from top to bottom and from left to right, or in habitual writing order of a nation.

TABLE 2

CODE	1	2	3	0
NAME	HORIZONTAL	VERTICAL	LEFT	RIGHT
SHAPE OF INDIVIDUAL STROKES				
DIRECTIONAL CHARACTERISTIC				

FIG. 2

GB 2 161 004 A

TABLE 1

CODE	1	2	3	4	5	6	7	0
NAME	HORIZONTAL	VERTICAL	LEFT-SWING	RIGHT-SWING	LEFT-OR CLOCK-WISE-TURN	RIGHT-OR ANTICLOCKWISE-TURN	CROSS	SQUARE
SHAPE OF INDIVIDUAL STROKES								
DIRECTIONAL CHARACTERISTIC								

FIG. 1.

TABLE 2

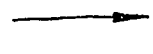
CODE	1	2	3	0
NAME	HORIZONTAL	VERTICAL	LEFT	RIGHT
SHAPE OF INDIVIDUAL STROKES				
DIRECTIONAL CHARACTERISTIC				

FIG. 2

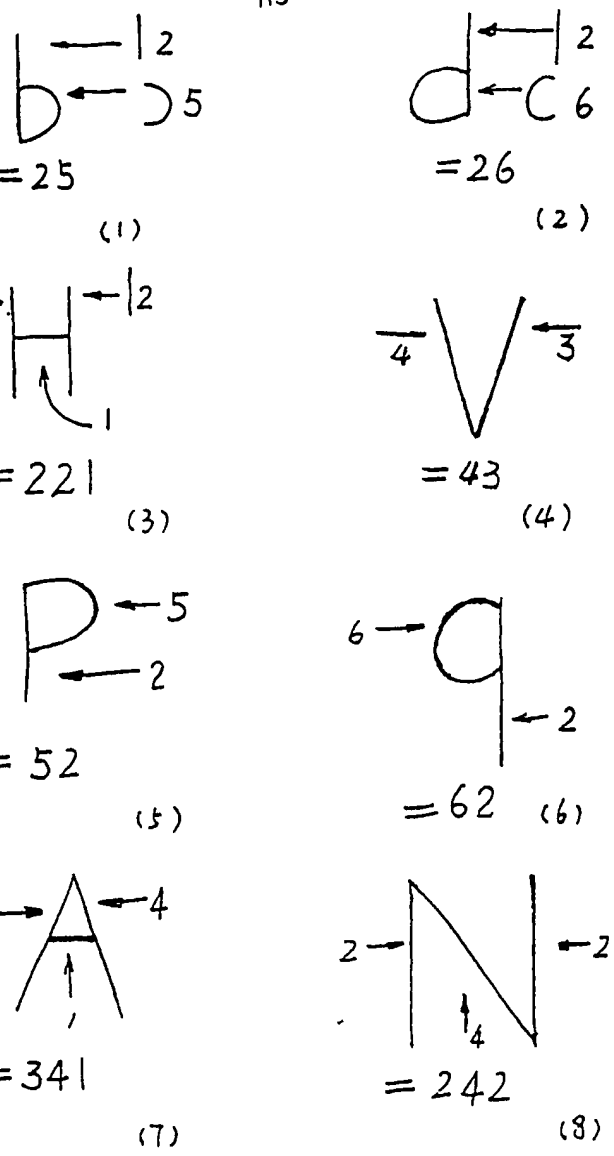


FIG. 3

ENGLISH LETTER	CODE	ENGLISH LETTER	CODE
O	0	A	341
Q	04	V	43
T	12	Y④	432(43*)
F	121	W⑤	434
J	15	P	52
Z①	16(6*)	R	524
E	161	B	525
I	2	C	6
L②	21(6**)	G	612
H	221	U	62
K	234	S	65
N	242	X	7
M③	243		
D	25		

FIG.4

ENGLISH LETTER	CODE	ENGLISH LETTER	CODE
O	0	w	434
q	02	j	45
g	05	t	51
e	06	p	52
z	16	r	55
l	2	c	6
h①	226(26*)	a	60
k	234	f	61
b	25	s	65
m	255	u	66
h	26	x	7
d②	260(26**)		
i	42		
v	43		
y	433(43*)		

FIG. 5

FRENCH LETTER (COMPLEMENTARY)	CODE
à	460
â	560
ç	63
ë	4406
é	306
è	406
ê	506
ï	442
î	52
ô	50
œ	006
ü	4462
ù	462
û	562

GERMAN LETTER (COMPLEMENTARY)	CODE
ö	440
ä	4460
ü	4462
ß	525

ROMANIAN LETTER (COMPLEMENTARY)	CODE
ţ	513
ş	653
ă	660
î	52

FIG. 6

ITALIAN LETTER (COMPLEMENTARY)	CODE
à	460
é	306
è	406
í	32
ì	42
î	52
ó	30
ò	40
ú	362
ù	462

PORTUGUESE LETTER (COMPLEMENTARY)	CODE
ã	160
õ	10
á	360
à	460
â	560
ç	63
ë	4406
é	306
è	406
ê	506
ï	442
í	32
ó	30
ò	40
ô	50
ü	4462
ú	362

FIG. 7

SPANISH LETTER (COMPLEMENTARY)	CODE
á	360
é	306
í	32
ó	30
ü	4462
ú	362

ESPERANTO (COMPLEMENTARY)	CODE
ĉ	56
ĝ	500
ĥ	526
ĵ	55
ŝ	565
ŭ	662

NORWEGIAN LETTER (INC. DANISH LETTER) (COMPLEMENTARY)	CODE
å	460
æ	600
ø	30

SWEDISH LETTER (COMPLEMENTARY)	CODE
ä	4460
å	60
ö	440

FIG. 8

RUSSIAN LETTER	CODE	RUSSIAN LETTER	CODE
О	0	В	225
е	06	А	341
Т	12	М	3434
Б	125	Л	32
Ъ	16	У	43
ф	20	ё	4406
Ю	201	З	5(55*)
Н	221	э	51
ш	2221	Р	52
щ	2225	ж	526
ц	225	т	55
И	232	L	6
К	234	а	60
b	25	ч	62
n	26	Я	62310231
bl	252	И	6232
д	2525	б	65
		Х	7

FIG. 9

VIETNAMESE LETTER (COMPLEMENTARY)	CODE
á	360
à	460
â	560
ă	660
ã	160
đ	70
Đ	75
é	306
è	406
ê	506
í	32
ì	42
ó	30
ò	40
ô	50
õ	60

VIETNAMESE LETTER (COMPLEMENTARY)	CODE
ú	362
ù	462
ý	343
ạ	604
ả	5660(560*)
ẹ	064
ẽ	106
ẻ	5606(506*)
ị	424
Ĩ	12
ọ	04
ỏ	50(506**)
Ơ	03
ụ	624
ũ	162
ủ	5662
Ư	623
ỹ	143
ỷ	5643

FIG. 10

JAPANESE KANA	CODE	JAPANESE KANA	CODE
ロ	0	タ	354
ニ	11	ル	36
テ	113	ン	41
エ	121	ウ	425
ヘ	14	ヴ	43
ラ	15	シ	441
モ	161	ツ	443
リ	23	ミ	444
ト	24	セ	45
ワ	25	ネ	4521
ノ	3	フ	5
ナ	31	ユ	51
ウ	313	ヲ①	51*
イ	32	コ②	51**
チ	331	ヨ	511
ク①	35	オ	513

FIG. 11.

JAPANESE KANA	CODE
ホ	5134
ア	53
ヌ	54
マ①	54*
ス②	54**
ソ	6
ヒ	63
ム	64
ヤ	65
メ	7
キ	71
ユ	721
サ	73

SERBIAN (INCLOFFICIAL YUGOSLAVIAN LANGUAGE) (COMPLEMENTARY)	CODE
Ђ	120
ђ	70
Ћ	125
ћ	75
Џ	355
Њ	2216
Ь	2213
j	45

MONGOLIAN LETTER (COMPLEMENTARY)	CODE
Θ	01

FIG. 12.

CHINESE CHARACTER	CODE
口	0
一	1
十	7
日	01
人	34
古	70
三	111
文	417
木	734
认	4634
如	6130
时	01514
任	32371
扩	511413
林	734734

CHINESE CHARACTER	CODE
课	4607734
鬻	0152153
麒	41322572
麟	41322574

FIG. 13

SPECIFICATION

System for encoding characters

The subject matter of this invention relates to a unified and simplified system for encoding characters of various kinds in the world. By this system, a kind of universal system that can easily process information composed of all kinds of characters in the world can be built up with the help of character-information processors (including computers, teleprinters and so on).

Up to now, "special keyboards" are needed to process information composed of characters other than English words with which the usual "standard keyboard" is enough to cope. When multi-kind-character information needs processing at the same time, it is necessary to have additional huge "multi-kind-character keybaord". Nowadays very many kinds of characters exist in the world, whereas so far no method and equipment are available for processing information composed of all kinds of characters.

The inventor thinks that all kinds of characters in the world are formed unexceptionally by different strokes. Occidental words are formed by letters in linear order which are constructed by different strokes. Oriental characters, such as Chinese characters and Japanese kanas, are 2-D graphs which are also formed by various strokes from top to bottom and from left to right. Occidental words appears to differ enormously from Oriental ones. Nevertheless, viewing from the fact that letters are also of a stroke form and a 2-D graph, alphabetic writing is in fact all the same to Chinese characters.

A patent application named "System for Encoding Chinese Characters" has been previously filed, the application numbers are as follows:-

US Patent Application No. P321

UK Patent Application No. GB 2 100 899 A

West Germany Patent Application No. P3217307.5

Japan Patent Application No.

The invention is mainly characterized by regarding all kinds of character is in the world as formed by different strokes which are grouped according to their shapes and writing directions, and allocating to each category of stroke an Arabic number, then encoding the strokes in strictly formulated order.

Detail description of the invention:-

(1) *Stroke Shape Codes:-*

In this invention, strokes which form all kinds of characters in the world are classified according to their shapes and directions. One may choose 8, 4, 16 or 32 categories. Fig. 1 shows a code table for 8 categories, which describes the shapes, directions and names of the 8 categories, and to each of the eight a digit code is allotted.

If 4-category classification is preferable, left-swing and left(clockwise)-turn in Fig. 1 may be incorporated into one category, and so do the right-swing and right(anticlockwise)-turn. Besides, both cross and square may be divided into two kinds of strokes, namely, horizontal and vertical (or left and right). Thus a 8-category may be reduced to a 4-category, of which each group is allotted a number code, i.e. 1, 2, 3, 0 (ref. Fig. 2), finally only horizontal, vertical, left and right four kinds of strokes are left.

If 16-category or 32-category classification is chosen, sub-divide left(clockwise)-turn, right-(anticlockwise)-turn, cross and square to get the desired classification.

4-, 8-, 16-, or 32-category classification is preferable to 5-, 10-, 20-category classification for reasons of easier adaptation to the binary system which is, by adopting such binary numbers as $4(=2^2)$, $8(=2^3)$, $16(=2^4)$, or $32(=2^5)$, etc., widely used in computers and other information processors for storing and processing information both conveniently and economically.

Generally, 8-category classification is most commonly used, referring to Fig. 1.

(2) *Rules for Encoding:-*

In this invention strokes are encoded by certain rules.

There are usually two rules, one is writing order rule and the other is positional order rule.

Writing order is also known as stroke order, i.e. habitual writing order of different kind of character. It varies with the person who writes and with the words which are written.

Positional order is one which is defined by the different positions of strokes in a character (or a letter). This invention regulates that the positional order is: first, the highest position, then, the next highest, up to the lowest, this is also known as from top to bottom; then, left side before right side, this is also known as from left to right.

When comparing two stroke positions, first observe the initial points of the two, as shown in Fig. 3(1), (2). If none is higher than the other, observe which one initiates from left side first, as shown in Fig. 3(3), (4). When the two initial points overlap each other, compare which termination point is higher, as shown in Fig. 3(5), (6). If the two terminal points have the same position, see which one is on the left side, as shown in Fig. 3(7), (8).

Although encoding in writing order may be consistent with the writing habit of a nation, ambiguity would arise due to th fact that this order varies with the person who writes and with the words which are written, thus resulting in a less strict and less unified encoding system. It may appear to be encoded.

Graphic order is a general rule refined from

writing order rule. It is more scientific and enables unification of the system for encoding all kinds of characters in the world. Meanwhile, it guarantees a unique meaning of each code. For this reason, graphic order is preferable in encoding and it results in both convenience and simplicity in use.

All kinds of characters in the world can be encoded by the above-mentioned method. Figs. 4–13 show how the encode English, French, German, Romanian, Italian, Partuguese, Spanish, Norwegian (and Danish), Esperanto, Swedish, Russian, Vietnamese, Japanese kanas, Serbian, Mogolian, and Chinese characters.

This system is also suitable to Korean Hangui, Arabic or any character of various national forms.

Of course, a character may be given two number codes wherever its graphic order differs from its writing order, i.e. two different codes refer to the same character and one more reference source could be used. Thus one can find out the exact character (or letter) by different rules.

(3) *Keyboard for Encoding:–*

The availing standard keyboard can be used as encoding keyboard for the present invention. One could also, however, have, say, 8, 4, 16 or 32 keys. The number of keys varies with how many categories of stroke shapes are adopted. Owing to the simplification and convenience of this system, it is possible to miniturize the complicated keyboard otherwise needed and greatly reduce the volume of ancillary apparatus.

Punctuation marks can also be encoded by grouping stroke shapes, e.g. colon ":" consists of two points, and its digit code is "44"; question mark "?" has a left(clockwise)-turn and a point, the code "54" stands for it, and so on, thus avoiding using special punctuation marks keyboard in the apparatus.

The great value of reduced keyboard lies in its enabling corresponding reduction of the computers which can be reduced to the size of a calculator and can be used in combination with a teleprinter and dial telephone to transmit information composed of all kinds of characters in the world.

The drawings:–

Figure 1, Table of 8-category system for encoding strokes shapes.

Figure 2, Table of 4-category system for encoding strokes shapes.

Figure 3, Graphic order rule based on 8-category system.

Figure 4, Digit codes for capital English letters.

Figure 5, Digit codes for small English letters.

Figure 6, Digit codes for complementary French, German and Romanian letters.

Figure 7, Digit codes for complementary Italian, Portuguese letters.

Figure 8, Digit codes for complementary Spanish, Norweigan (including Danish), Esperanto and Swedish letters.

Figure 9, Digit codes for Russian letters.

Figure 10, Digit codes for complementary Vietnamese letters.

Figure 11, Digit codes for Japense kanas.

Figure 12, Digit codes for complementary Japanese kanas and Serbian (official Yugoslavian language) letters.

Figure 13, Digit codes for Chinese characters.

Description of the Drawings:–

In Fig. 4;

(1) Z may be given a code "16" or "6". Since the latter is identical with the code for C, there will be a confusion if code "6" is used for the two. Therefore, an "*" mark is added to tell one from the other. Thereafter, "*" mark appears wherever a code stands for different characters.

(2) L may be given a code "21" or "6", for the above-said reason, it is encoded as 6**.

(3) M should have been encoded as "234", had the code "234" not differentiated from other codes. Since the code "243" is an effective code, drop the last numeral 2 which is useless.

(4) Y may be encoded as "433" or "43"; the latter is identical with the code for V, add "*" mark here.

(5) W should have been encoded as "4343"; since "434" is effective, drop the last number 3.

In Fig. 5:

(1) H may be encoded as "226" or "26". Since the letter N anticipates to take "26" as its code, the two letters would have same code if "*" mark would not be added.

(2) D may be encoded as "260" or "26", add "*" mark to "26".

In Fig. 9:

3 may be encoded as "5" or "55". The latter has been taken, thus add "*" to it.

In Fig. 10:

(1) 2 may be encoded as "5660" or "560"; the latter is identical with other code, add "*" mark here.

(2) 2 may be encoded as "5606" or "506"; the latter is identical with other code, add "*" mark here.

(3) 8 may be encoded as "50" or "560"; the latter is identical with other code add "*" mark here.

In Fig. 11:

(1) ㄱ may be encoded as "51" which is identical with other codes; add "*" mark here.

(2) ㄷ may be encoded as "51" which is identical with other codes; add "**" mark here.

In Fig. 12;

(1) ㄱ may be encoded as "54" which is identical with other codes; add "*" mark here.

(2) ㅈ may be encoded as "54" which is identical with other codes; add "**" mark

here.

CLAIMS

1. A method of encoding all kinds of characters in the world in digits, in which the key point is to find out individual stroke in 2-D direction of all types of characters the world over, including Occidental alphabetic writing in 1-D linear arrangement and Oriental ideographic characters in 2-D arrangement (such as Chinese characters), and allocate a digit to each specific category of stroke shape, thus making it possible to encode all kinds of characters in the world systematically.

2. A method as claimed in Claim 1 in which the strokes of all kinds of characters in the world which are grouped into 8 basic categories of stroke shapes according to their respective characteristics and are allotted digit codes 1, 2, 3, 4, 5, 6, 7, 0 are horizontal, vertical, left-swing, right-swing, left(clockwise)-turn, right(anticlockwise)-turn, cross and square (referring to Fig. 1).

3. A method as claimed in Claim 1 in which the strokes of all kinds of characters in the world which are grouped into 4 basic categories of stroke shapes according to their respective characteristics and are allotted digit codes 1, 2, 3, 0 are horizontal, vertical, left (including left-swing and left(clockwise)-turn) and right (including right-swing and right(anti-clockwise)-turn).

4. A method as claimed in Claim 1 in which the strokes of all kinds of characters in the world can be grouped into 16 or 32 basic categories of stroke shapes according to their respective characteristics and be allotted digit codes accordingly.

5. A method as claimed in all preceding Claims in which the stroke shapes are encoded in graphic (positional) order, namely, first, the highest position, then, the next highest position, up to the lowest, this is also known as from top to bottom; then, left side before right side, this is also known as from left to right.

6. A method as claimed in Claims 1–4 which the stroke shapes are encoded in writing order, namely, characters of various nations are encoded according to their respective writing habit.

7. A method of claimed in any of the preceding Claims in which a letter or simple ideographic character is allotted 1–3 digits, ordinary ideographic character is alloted 6 digits and no more than 8 digits are used for most complicated ideographic character.

8. A method as claimed in Claims 1–6 in which all kinds of punctuation marks can be encoded by the system for encoding stroke shapes.

9. All kinds of dictionaries, codes and character indexes in which characters of various nations are encoded systematically by a method as claimed in any of the preceding Claims.

10. All systems for processing character information (including computers, teleprinters, typewriters, dial telephones, etc.), irrespective of large or super-mini apparatus, in which all kinds of characters in the world are encoded systematically by a method as claimed in any of the preceding Claims, one-machine-system which can process one, two, three, up to all kinds of characters in the world by a method as claimed in any of the preceding Claims.

Printed in the United Kingdom for
Her Majesty's Stationery Office, Dd 8818935, 1986, 4235.
Published at The Patent Office, 25 Southampton Buildings,
London, WC2A 1AY, from which copies may be obtained.